Serious Gaming for the Evaluation of Market Mechanisms

Fabian Lang  
Albert Ludwigs Universitat Freiburg, fabian.lang@is.uni-freiburg.de

Tim Pueschel  
Albert Ludwigs Universitat Freiburg, tim.pueschel@is.uni-freiburg.de

Dirk Neumann  
Albert Ludwigs Universitat Freiburg, dirk.neumann@is.uni-freiburg.de

Follow this and additional works at: http://aisel.aisnet.org/icis2009

Recommended Citation  
http://aisel.aisnet.org/icis2009/97
SERIOUS GAMING FOR THE
EVALUATION OF MARKET MECHANISMS

Research-in-Progress

Fabian Lang  
Chair for Information Systems Research  
Albert-Ludwigs-Universität Freiburg  
Platz der Alten Synagoge,  
79085 Freiburg, Germany  
fabian.lang@is.uni-freiburg.de

Tim Püschel  
Chair for Information Systems Research  
Albert-Ludwigs-Universität Freiburg  
Platz der Alten Synagoge,  
79085 Freiburg, Germany  
tim.pueschel@is.uni-freiburg.de

Dirk Neumann  
Chair for Information Systems Research  
Albert-Ludwigs-Universität Freiburg  
Platz der Alten Synagoge,  
79085 Freiburg, Germany  
dirk.neumann@is.uni-freiburg.de

Abstract

Design science consists of two major design processes: building and evaluation. A well-executed evaluation of design artifacts is crucial to their success. Traditional evaluation tools have certain weaknesses because design artifacts include “wicked” problems. Serious Gaming can help to overcome these problems. To this end an online based cloud resource managing game is developed which simulates the implementation of a market mechanism and represents a new design artifact. This mechanism is a heuristic solution consisting of dynamic pricing and a priority policy. The aim of this research is to show that Serious Gaming complements traditional evaluation tools and improves the evaluation of market mechanisms. Therefore, a general guideline for designing Serious Games for evaluation is developed and a classification of Serious Gaming is established. After having collected sufficient data, future work will be to analyze players’ behavior and finally evaluate the market mechanism.

Keywords: Serious Gaming, Design Science, Market Engineering, Evaluation Methods
Introduction

In the early 1890s, the Belgian chemist Leo Baekeland invented an improved photographic paper which did not need sunlight for developing pictures. A few years later Baekeland sold his patent rights to the Eastman Kodak Company for $1 million (Amato 1999). Initially, Baekeland planned to sell the rights for a sum somewhere between $25,000 and $50,000. Since Eastman bided before Baekeland could reveal his asking price, Baekeland brought off a big deal. This shows us that market design can really change the outcome of a market (Weinhardt et al. 2003). Economic science tries to give an insight into how markets work, Karl Marx however said: “The philosophers have only interpreted the world in various ways - the point however is to change it” (Marx 1845). This proposes a different view of economics: economics as a design science (Varian 2002). The design science consists of two major design processes: building and evaluation (March and Smith 1995). The latter is very important for the success of designs (Hevner et al. 2004). This paper focuses on the evaluation aspect in terms of searching for alternative evaluation tools for design artifacts.

Hevner et al. (2004) define artifacts as constructs, models, methods and instantiations applied in the development and usage of information systems.

“Artifacts are innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, and use of information systems can be effectively and efficiently accomplished.” (Hevner et al. 2004)

The examined design artifact is a policy-based heurist for resource allocation in cloud computing. In the case study, this market mechanism is going to be evaluated by comparing the results of the mechanism with the outcome of human test subjects in a game which depicts managers’ allocation decisions.

Design implementation is a “wicked” problem since it is a situation in which one cannot gain experience from this situation because of the situation’s uniqueness. Furthermore, the implementer is threatened by consequences which cannot be revoked. Like an architect constructing a bridge, a market designer has a single attempt for the implementation of just one blue print. If the architect fails, the bridge will collapse, just like a provider will not earn the best possible revenue if the allocation mechanism is not appropriate (Rittel and Webber 1973). Hence a pathway is required in which an artifact can be tested without risking real consequences. The outcome of market engineering is determined by agent behavior and the market structure (Weinhardt et al. 2003). A better market outcome can be anticipated by utilization of improved evaluation tools. Usually market designers use simulations, pilot runs, analytical approaches or laboratory experiments for analyzing agents’ behavior and design evaluation before implementing mechanisms (Neumann 2004). However, these conventional evaluation tools have several weaknesses which will be discussed in section “Traditional Evaluation Techniques”. The paper proposes that research has to take a step from simulated behavior in simulations to real behavior in a simulated environment. This environment can be implemented within a serious game. This approach can be favorable since the agents’ behavior is real and not distorted by assumptions about their behavior. Assumptions about the market environment are needed for a serious game and they are needed for a simulation as well. Using games as a simulation is not a new idea. Military officers have been using war games in order to train strategic skills for a long time. One of the oldest games is the Indian board game called Chaturanga or Wei hei, which was played approximately 2,000 BC (Susi et al. 2007). Another example for early serious gaming is an early 19th century Prussian military training game called “Kriegsspiel” (German for wargame), which included multiplayer playing (up to 10 people), communication and information sets (von Hilgers 2000). Even in the gulf war in 1990 the Pentagon used war board games to figure out strategies.

Serious Gaming (SG) as an evaluation tool is a relatively unexplored scope in economic research. We want to show that serious games (SGs) are an efficient way of evaluating artifacts.

Following this brief introduction, we are going to discuss the main evaluation tools and SG as such a tool. Finally we are going to introduce the game setup for the case study and show our outline for the further progress of our research.
What is a Serious Game?

The term “Serious Games” (SGs) is an oxymoron, a composition of two contradictory terms. Games are usually not associated with a productive way of spending time. However, SGs want to combine a higher purpose with an entertaining environment (Michael and Chen 2006; Zyda 2005). This concept is already used in education and training games (in schools, companies or the military), advergames (advertisement embedded in games), health games (e.g. a driving simulation for people with driving phobias) or games as work (like professional online casino player) (Dahl et al. 2009; Stapleton 2004). SGs are a small but rising fraction on the $10 billion video game market of the US, representing a $20 million industry in the US (approx. $70 million worldwide) (Sawyer 2007; Susi et al. 2007; van Eck 2006).

For this research paper the following definition is used, based upon Zyda (2005):

A Serious Game is a mental contest which is played in accordance with specific rules to observe a certain result or behavior of the players.

Traditional Evaluation Techniques

Analytic Approaches

Theoretical models or game theoretical approaches are an abstract way of evaluating artifacts. The center of theoretical models is a formal representation which is associated with very restrictive assumptions and constrains the analysis. The divergence between model environment and real world has to be considered. Therefore, theoretical models as stand-alone indicators are not sufficient for the evaluation process. Aggregate theoretical views are mostly dominated by agent-based techniques. Theoretical models’ strength is the ability to give an economical intuition of the problem and to separate between important and unimportant factors (Neumann 2004; Schieritz and Milling 2007). Game theory studies the behavior of rational agents who react to the behavior of other agents. It is an easy tool to analyze ordinary problems. The evaluator has to handle multiple decision factors, inconsistent behavior, interference of an outside party, time-lags, uncertainties and conflicting goals (Babb et al. 1966; Korhonen et al. 1986). Numerical computation takes too much time caused by big strategy spaces. Although computation takes long, it does not yield a feasible solution every time (Neumann 2004). Game theory will not deliver exact solutions but the results can be seen as useful approximations. The outcomes might be weaker than expected (Milgrom 2004).

SG introduces real agents to the evaluation. While analytic approaches try to incorporate every factor of the issue, real agents act heuristic and pragmatic. Adopting artifacts, which are only examined by analytical approaches, can lead to unexpected outcomes because people act rather reasonable than rational (Rawls 1993).

Simulations

Simulations are defined as “the imitation of the operation of a real world process or system over time” (Banks 1998). Simulations can provide a significant and controllable environment which can deliver an answer to various scenarios. Besides, simulations can predict behavior in easy problem sets surprisingly well. Also, they can help managers developing a better understanding of processes and operations (Banks 1998).

Nevertheless, there are also a few disadvantages: Simulations need a lot of resources for calculation and a big input of time and knowledge is required. Moreover simulation technology costs are relatively high including hardware, integration, software licenses, maintenance of technology and training costs (McLean and Leong 2001). These restrictions often lead to a fragmentarily conducted analysis. Due to restricted time and knowledge, professionals need heuristic solutions (Law 1983). Furthermore, the evaluation of a market mechanism is a “wicked” problem. A “wicked” problem has certain characteristics: it lacks definitive formulation, stopping rule and ultimate test. Simulations cannot handle these features properly (Rittel and Webber 1973). They also bear some methodical weak points like using just a given set of input parameter or severe theoretical assumptions in model building (Schnizler 2007). Furthermore, simulations can suffer from an absence of useful benchmarks.

SG can serve as a benchmark and represent pragmatic human intelligence. Hence SG can complement traditional simulations. SGs are associated with high technology and human resource costs as well. But in contrast, games need no knowledge about behavior or the dimension of the issue because this will be explored by the players. Strategies which are not mentioned by the professionals can be found as SG utilizes collective intelligence. An example that SGs can outperform simulations: At the very beginning of the gulf war, the Pentagon wanted to figure out an optimal response strategy. Unfortunately, simulations could not answer this question sufficiently. The Pentagon decided to make use of the commercial war board game “Gulf Strike” to simulate and negotiate most of the later real strategies (Dunnigan 1992; von Hilgers 2000).
Field Studies and Laboratory Experiments

Field studies and test runs implement artifacts in a separated sphere e.g. a specific regional market. Laboratory experiments implement these artifacts in a test environment under well-defined conditions. Field studies have the advantage that they are conducted in a real environment. However, they are expensive, can cause negative consequence, like losing customers in the test market and cannot evaluate a variety of alternatives. Moreover, it is difficult to control all experiment factors in the field and reproducibility is not guaranteed.

Laboratory experiments are much cheaper than field studies and do not contain real risks. Furthermore, they can examine different artifacts alternatives very easily. However experimentalists need a sophisticated setup of experiment and incentives for subjects, otherwise experiments can draw wrong conclusions (Muller 1999).

SGs have a key incentive for participating: entertainment for the participants. Subjects are not as aware of being observed as if they would be invited to attend a test in a laboratory. They do not feel forced to answer in a manner which the investigator wants to hear (observer-expectancy effect). An investigator is not required for SGs and a great number of test persons can be acquired. Laboratory experiments are often biased through having almost exclusively students as test persons. SG can be open to everybody but it can be excludable as well. Professionals are usually hard to obtain for an experiment, however SG is easy to use and neither location nor time bound. So it can facilitate obtaining professionals. A further disadvantage of laboratory experiments is the Hawthorne effect. Test persons in SGs do not have to be aware of taking part in a study as long as they are made belief that it is an ordinary game. The research purpose can stay undisclosed. This can prevent hypothesis guessing as well.

Summarizing, SG can improve the projection of real world using real agents. A combination of SGs, experiments and simulations is conceivable and beneficial for evaluation.

Serious Gaming as an Evaluation Tool

Evaluation Market Mechanisms with Serious Gaming

Classification of the Method

In accordance with Hevner et al. (2004), the classification of SG as an evaluation method takes place within two dimensions: formalization level and paradigm. Formalization represents the distinction between qualitative and quantitative methods. Paradigm is divided into behavioral and constructive approaches. Behavioral methods are focused on impacts on companies, people and markets. Constructive methods are mainly interested in the development of artifacts.

For classification of SG see figure 1 which depicts paradigm on abscissa, formalization on ordinate2 and frequency of usage in the area of the circles. Frequency was adopted with data by Palvia et al. (2004) who conducted an empirical analysis of research methods in famous information systems research journals3 between 1998 and 2003. SG is classified as a behavioral quantitative method (denoted by the gamepad in figure 1). This results from the fact that SG has a formalization level equal to reference models and a similar paradigm like laboratory experiments. Reference models try to simplify circumstances formally and represent business situations understandably. They are deducted inductively or deductively from empiricism or theoretic models and try to develop optimal concepts for information systems (Wilde and Hess 2007). Since game developers have to model the business environment, SG uses simplified concepts. Therefore, SG takes place in a quantitative environment.

Because it tests artifacts under conditions comparable to real implementation, SG is very close to the idea of laboratory experiments. The methods attempt to find conclusions from test persons’ behavior through observation. Therefore, SG is a behavioral approach with constructive influence through the simulated implementation. SG is not congruent with laboratory experiments as mentioned in the section “Traditional Evaluation Techniques”.

---

1 As test persons feel privileged to be a part of an experiment, they tend to change their behavior.
2 Characteristic of dimensions are taken from Wilde and Hess (2007), except method of SG.
At first glance, there might be inconsistencies by proposing a behavioral method as a design evaluation tool for constructive design science. But this criticism can be neglected: Behavioral science’s purpose is to find explanations of phenomena, to find the truth, whereas design science’s objective is to find solutions for problems and yield utility. Both are inseparable. A certain quantum of theory is needed for design whereas construction can result in a new chapter of unexplored theory (Hevner et al. 2004). An architect cannot construct a bridge without knowledge of statics, materials or physics. Analogously, an economist has to know about customers’ preference, technology acceptance and behavior when designing a market mechanism. Referring to Karl Marx’ quote from the beginning, philosophers must interpret the world in order to change it.

Evaluation Design

The evaluation is a critical part of the design process. We have to create a business environment for the evaluation which is a new design artifact in its own right. The purpose of the evaluation is to reveal the mechanism’s quality, efficacy, efficiency, adaptability and utility. To accomplish this purpose, the evaluation has to be executed carefully and rigorously. If the prior defined requirements are satisfied, the evaluation either results in suggestions for improving the mechanism or authorizes the actual implementation. Considering that design is a “wicked” problem, evaluated criteria must be defined due to relevance and rigor (Hevner et al. 2004; Rittel and Webber 1973).

A brief guideline for a SG evaluation is given in figure 2 orientating itself to George Pólya’s sequence of heuristic solutions (Pólya 1945).

---

4 Maybe that is why his state ideology failed in the “field trial”. 
(1) problem definition

As mentioned above, the problem and the purpose have to be defined at first. This information is important for step 2. Based upon the problem definition, the Serious Game has to be constructed.

(2) game conception

The game conception features two main factors: game design and rule design. Game design represents our economic environment, whereas rule design is important for interaction between players or the game’s sequence. For instance, the rules of the MIT “beer distribution game” prohibit communication between value chain groups. This changes the outcome of the game significantly (Sterman 1989). Test runs are important to find out if the game can satisfy the requirements and user-friendliness. Test runs are also an evaluation method and are used for evaluation of the game which represents a new created artifact in its own right. The evaluation of the game is needed to prevent errors in the conception.

(3) playing

In step 3, the game is implemented and test persons can play the game. If expert knowledge is required, well-informed players have to be found. Players should be given the opportunity to play test rounds and to play the game repeatedly. This can prevent playful manner and accustom the player to the game. Competition and award can also improve results because players participate more actively and ambitiously (Babb et al. 1966; Binmore 2007).

(4) analysis

In the analysis, the mechanism’s result is set in proportion to the players’ results. Comparing the results to a further benchmark can improve validity. Benchmarks can be an automatism or a computed theoretical solution, for example. Further observations of effective strategies and successful behavior should be used to improve the artifacts. Alternative game settings can be introduced in order to compare results on different suppositions. Additionally, the players can be confronted with shocks and crises. The reaction of the players and the mechanism in abnormal situations is interesting for further research projects.

Case Study: Online Based Cloud Managing Game

A Market Mechanism for Clouds

The mechanism is a policy-based heuristic for resource management in cloud computing. It is taken from Püschel and Neumann (2009). A dynamic pricing policy is proposed which takes into account that capacities are limited. It adjusts ask prices as opportunity costs for reducing capacities. In this scenario, we assume that there is a client classification including “gold clients”, whose requests are given priority. The heuristic is formalized as follows:

$$\max_\mathbf{x} \sum_{j=1}^{J} \left( \frac{1}{2} + x_j \right)^k$$

subject to

$$\sum_{j=1}^{J} c_{rp}(t) + p_j \leq c_p(t)$$

$$\left(1 - \Theta(p_j - p_{rp}) \right) \times \frac{\sum_{j=1}^{J} c_{rp}(t) + p_j \leq c_p(t)}{c_p(t)} \leq 1_{p_{rp}}$$

$$\left(1 - \Theta(p_j - p_{rp}) \right) \times \frac{\sum_{j=1}^{J} c_{rp}(t) + p_j \leq c_p(t)}{c_p(t)} \leq 1_{p_{rp}}$$

with

- \( I \): requests
- \( J \): resource
- \( x_j \): allocation variable (binary)
- \( c_{rp}(t) \): resource requirement
- \( p_j \): bid price
- \( \Theta \): Heaviside step function
- \( \Theta \): thresholds
- \( \rho \): period

The objective function encodes information about the ascending order of requests and the capacity constraint implies that utilization must not exceed capacity limits in any period. Introducing different utilization stages, the ask price can be calculated as a step function of the utilization. Finally, the mechanism suggests a higher ask price for non-gold client requests, if utilization exceeds a certain percentage of the available capacity. The consequence of treating clients with different priorities is said to provide system stability and Quality of Service (QoS).
The Cloud Managing Game

(1) problem definition

The objective is a performance evaluation of the market mechanism. The problem set is given through the mechanism and the game structure. The performance of the players has to be compared with the mechanism. Furthermore, the behavior of players has to be analyzed. SG serves as a benchmark. Additional benchmarks are a policy which adopts “first-come, first-serve” and the theoretical optimum. Both benchmarks are not as suitable as SG since “first-come, first-serve” is a dull policy and the theoretical optimum is not feasible due to lacking information and uncertainty.

(2) game conception

The game is set in a cloud computing firm. The player as the economic agent (EA) has to manage resource allocation in order to maximize the revenue. The game is online based and is designed with HTML, php and JavaScript, including a MySQL database.\(^5\)

![Figure 3. Screenshot of the game design](http://is-games.vwl.uni-freiburg.de/cloudmanager)

The setup of the scenario contains three resource types \( r = \{\text{processing power, bandwidth, memory}\} \) with a fixed capacity of \( c_r \) and three products \( s = \{\text{calculation (0.16/0.04/0.08), backup (0.04/0.04/0.16), video hosting (0.04/0.16/0.08)}\} \), where numbers in round brackets denote the usage of resource \( r \) by one unit of service \( s \) in percent of the capacity. One single request covers several instances and resource demand for one request is pretty high (up to 16%). These simplifications are needed to make the issue cognizable to the players. In real cloud computing several requests are submitted per second making human real time decision impossible. Slowing decision speed down does not affect validity of human decision. Ten job lists are created in advance. These jobs can start off either right after submission or in a future period. The runtime of the jobs varies. Job requests also contain information about the number of instances asked and the bid price for this job. The EA can cancel running tasks but has to pay a contractual penalty which depends on the full price. There are 25 imaginary customers; five of them are gold clients. In order to secure Quality of Service and system stability, overload situations are prohibited. If capacity would be exceeded otherwise, the EA has to cancel a running task or reject a request (Püschel et al. 2007). If a gold client request is rejected, the player has to pay a penalty representing a lost benefit caused by the violation of the gold client priority. After a given amount of periods the game ends and all running tasks are paid out to keep efforts stable.

(3) playing

In order to get large data and an economic educated group, graduate economics students as well as doctoral students are chosen as players. Players can play the game repeatedly, so they can apply both playful and serious strategies. Besides being mentioned in a high score list, the best player can win a gift coupon if s/he has outperformed the market mechanism. This should be a sufficient incentive for participating and acting seriously.

---

5 The game is accessible under http://is-games.vwl.uni-freiburg.de/cloudmanager
A calculation of the theoretical optimum under perfect information and certainty is made. Each player’s and the mechanism’s performance is measured by dividing the revenue through the optimal revenue yielding a percentage ratio of the performance. The theoretical optimum is computed under the assumption of full information hence this solution is not feasible due to unavailable information about future jobs. A further benchmark is introduced realizing a “first-come, first-serve” policy. This policy lacks sophistication and therefore it is not very convincing. That is why SG is needed. The proficiency of the market mechanism is concluded out of the performance of the average player and the high score players. Afterwards, the mechanism is adjusted adopting strategies used by successful players.

**Preliminary Results**

After the first test rounds, it was exposed that the mechanism beats human players on average. Single players can outperform the mechanism, however not consistently. Interviewing the best players yielded that they used some kind of ask prices like the mechanism. If the player had an information advantage similar to the theoretical optimum (e.g. unpunished cancelations of running jobs), the best players would have achieved a performance close to the optimum proving that human being can handle such problems. Since there were no available mechanisms for comparison, SG appeared to be a good benchmark.

**Conclusion and Outlook**

Complex problems can be too complicated for laboratories. SGs can illustrate very vague coherences without confusing the player since the coherences are hidden in the game engine. While the contribution in a laboratory experiment is mandatory when test persons signed up for it, SGs have not any commitments since they can be played whenever the player chooses to.

Organizational issues are crucial to the success of the development and the implementation of a design artifact (Hevner et al. 2004). Games incorporate the economic modeling of simulations, but also provide much more opportunities for the setting. For instance, lagging and lacking information can be included. Therefore, SGs have a high potential on studying organizational issues. Unlike game theoretical approaches, where people try to understand behavior by considering mostly rational agents and severe assumptions, SG can simulate the same situation without these hard limitations and with a more detailed reality. Game designers can vary communication opportunities, the quantity of decision variables, ways of interaction between players, time delays, crisis vs. non-crisis decisions and uncertainty very easily (Barone et al. 1975). Scientists do not need severe assumptions, but only a set of rules implemented in the game design. Babb et al. (1966) showed that played policies are congruent with real-life decisions; therefore we can measure behavior using a business simulation.

Human beings have another way of accessing information than computers: They have restricted memory capacities and awareness and make use of heuristics. Furthermore, they cannot build expectations properly and are driven partially by emotions (von Nitzsch 2002). Consequently, SG uses heterogeneous agent models and is very beneficial for market interaction problems.

If the market mechanism is evaluated as being weak due to the players’ superiority, the design of the mechanism can be improved by analyzing players’ manners. If the mechanism can outperform the players, this is no reason to stop improving the mechanism. Depending on the setting, people have limited capabilities to comprehend contexts and solve the setting’s problems. Nevertheless, they can represent a benchmark. This is valuable all the more if other benchmarks lack validity like in the case study.

The preliminary results showed that SG is suitable for an evaluation. Furthermore, Hevner et al. (2004) stated that comparison of an evaluation method with alternative approaches is crucial for claims of generalization. SG is such an alternative approach.

Concluding, the study is going to show if the suggested market mechanism performs well by using SG. SG is a promising research method which will be supported through further research. The on-going research is going to be a more detailed demonstration of the potential of SG. The market mechanism represents a heuristic solution to resource management and can improve provider’s revenue and customer’s QoS-level. Until now the scope of SG is not explored very well. We hope that research is going to come upon new findings.
References


Sawyer, B. "What We know About Games in Classrooms Now," Instructional and Research Technologies Symposium for the Arts, Humanities and Social Sciences, Rutgers, The State University of New Jersey, 2007.


